63-3-2

101

TM-1150 000 00

A Study in Probabilistic Information Processing (PIP)

TECHNICAL MEMORANDUM

(TM Series)

ASTIA AVAILABILITY NOTICE

Qualified requesters may obtain copies of this report from ASTIA.

This document was produced by SDC in performance of contract SD 97 (ARPA)

A Study in Probabilistic Information Processing (PIP)

by

Richard J. Kaplan & J. Robert Newman

2 April 1963

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA

SUMMARY

Men required to choose among alternative hypotheses given fallible data fail to extract as much "certainty" as the data justify. PIP Theory indicates that performance can be improved by using Bayesian probability judgments. This experiment was designed to test this theory.

Subjects, receiving simulated military data, determined which one of three strategies (Hypotheses) an enemy was using. In the NON-PIP condition subjects estimated P(H/D) directly. In the PIP condition subjects estimated P(D/H), and P(H/D) was calculated using Bayes Formula. Results show: (1) the highest probability was always assigned to the correct hypothesis. (2) PIP was superior to NON-PIP in (a) achieving higher posterior probabilities, and (b) reaching asymptotes faster. (3) Increasing difficulty resulted in poorer performance in NON-PIP.

INTRODUCTION

All systems of men and machines deal with fallible data and use such data to make decisions and carry out their missions. There are a large number of ways of processing such data to refine them, improve them, and hopefully make them more utilizable. The field of statistics has offered many methods and techniques to make inferences from such data, and recently a class of procedures referred to as Bayesian statistics have been brought to bear on these problems. It is within the Bayesian framework that the structure of this experimentation has been built.

Basic Theory

Edwards (1962) has specifically proposed a class of information processing systems that capitalize on the Bayesian techniques. These he has called Probabilistic Information Processing (PIP) systems. The principal objective of a PIP system is to get maximum use of fallible, uncertain data in making diagnostic decisions, i.e., decisions about which one of several hypotheses concerning the state of the world is correct. The mathematical basis of a PIP system is Bayes's theorem, which is a conditional probability statement of the following kind:

> Given a set of exhaustive and mutually exclusive hypotheses H_i, and a set of events D which are known to have occurred, the probability that it was H_i that resulted in D is
>
> $$P(H_i/D) = \frac{P(H_i)\, P(D/H_i)}{P(D)}$$

The denominator $P(D) = \sum_{i}^{n} P(H_i)\, P(D/H_i)$ and, of course, must be non-zero.

The $P(H_i)$ are the a priori probabilities of H_i, $i=1, 2 \ldots, n$. They must be available initially and are usually estimated by whatever procedure is appropriate. The $P(H_i)$'s are commonly assigned equal probabilities, and it has been found by Edwards that the initial values have little effect on subsequent system operation as long as they are not too close to either 0 or 1. For example, if there are three hypotheses, each would be assigned the initial probability of 1/3. Thereafter, the process is sequential and cumulative; the output of one Bayesian calculation, $P(H_i/D)$, for datum-event k is used as an input for $P(H_i)$ for datum k+1. The denominator of the above equation is essentially a normalizing factor which assures adherence to the probability axiom that

$$\sum_{i=1}^{n} P(H_i/D) = 1.$$

The probability $P(D/H)$ is not readily available or easily estimated. It is the probability of observing the datum D if H_i were the true hypothesis. Edwards has suggested that humans may be able to estimate $P(D/H)$. In other words, we would use humans essentially as probability estimators to provide a $P(D/H)$ for each datum and each hypothesis. These probability estimates are fed to a computer, which calculates the $P(H/D)$'s using Bayes's theorem. This process is repeated sequentially. The resultant output of a PIP system is a continually updated representation of the system's opinion as to which of the hypotheses is correct.

A PIP system relies heavily on human judgment, and the probabilities used in the Bayesian processor are actually defined as orderly opinions on the part of the decision maker. What Bayes's theorem does is to show how the evidence of observations (data) should modify or confirm previously held beliefs in the formation of expert opinion, and how, on the basis of such opinions and value judgments, a choice can be made between alternative hypotheses. The role of

human judgment, however, is quite different from that commonly employed by present-day systems. In a PIP system the human is being asked to evaluate the probability of a datum given some hypothesis. The burden of calculating the likelihood of the hypotheses being correct is carried entirely by the Bayesian processor in the computer.

In this paper we do not present the rationale and theory behind a PIP system. This has already been done by Edwards (1962). It should be noted, however, that there are two fundamental assumptions that form the underpinnings of the theory. These are: (a) Men can be trained easily to handle uncertain data and to perceive whatever probabilistic structure is inherent in such data. Initial evidence also indicates that the probability estimates supplied by PIP operators merely have to be consistent, not necessarily accurate. (b) Men are not as good in reaching conclusions about the significance of such data, especially in deciding what hypothesis is best supported by the data. Men are apparently quite conservative and are unable to squeeze as much "certainty" out of the data as is actually there. A PIP system capitalizes on man's ability to estimate probability but automates the final conclusion. There is some evidence to support both of these assumptions, but considerable empirical research is needed to bolster this support. Fortunately, both assumptions are directly testable.

To give an example, consider medical diagnosis of a patient who might have a certain type of cancer. The examining physician using a PIP system would consider each diagnostic sign or symptom with reference to each of the possible hypothesis, i.e., types of cancer, and would assign a probability to the observation of that sign or symptom when given each of the alternative hypotheses. These would be inserted into a Bayesian Processor and the probability of each

hypothesis, given that sign or symptom, would be calculated. The process is then repeated for each additional diagnostic sign. In a non-PIP system the physician would consider all the diagnostic signs and symptoms together and give his best estimate (probabilistically) of which hypothesis would be correct. There are many military decisions problems that parallel this example of medical diagnosis. For research purposes, we have abstracted one of these military problems to use as a "test bed" for investigating PIP principles. The specific military problem selected is that of recognizing and interpreting the significance of a nuclear detonation that has occurred somewhere in the United States. This is the mission of the NUDETS system, the essence of which has been abstracted for experimental purposes.

METHOD AND PROCEDURE

Summary

This experiment used an abstracted NUDETS data base in which a subject was given a "report" indicating the location of a particular nuclear detonation. He was told that the enemy intended to bomb either a military, civilian, or industrial target and aimed the nuclear bomb to hit one and only one of such targets. The subject was also told that there was a certain amount of error in any type of bombing, and that the likelihood of the enemy actually hitting the target at which he had aimed was never certain. The subject's basic task was to decide which target the enemy had selected. Thus, if each target was considered to be an a priori hypothesis of enemy intent, the problem was to decide which hypothesis was correct. In the non-PIP case the subject did this directly, i.e., he considered each datum and its referent geographical location of nuclear detonation and decided whether the true target class being attacked was military, civilian, or industrial. He then assigned a probability to this hypothesis which indicated the

degree of belief he had in its correctness. In the PIP case, the subject's task was to consider each datum in reference to each hypothesis independently and assign a probability to that datum. This number was fed to the Bayesian Processor and $P(H_i/D)$ is calculated using Bayes's theorem.

Basic Task

The experimental task required that the subjects perform an analysis of data from a simulated and somewhat abstracted NUDETS system. The stimulus material for each subject for a particular session consisted of a forty-page booklet of target maps, NUDETS report for each target map, and sheets on which to record his responses. The NUDETS report consisted of the X, Y, coordinates of a hypothetical bomb burst printed on a 3" x 5" card opposite the proper map page of the target booklet. The subject was required to locate the bomb burst on the map-area grid and to make the appropriate probability estimate. He then recorded his estimate on the response sheet, and went on to the next page in the booklet. At the completion of one booklet, the subject handed it to the experimenter and was given the next one.

The Target-System Stimulus

The target system comprising the subject matter for the experiment was a hypothetical rectangle 500 x 800 miles in size. This area was divided into 40 squares, 100 miles on a side, and each square was represented by one page of the target booklet. Grid lines of 10-mile divisions were indicated on the page and the squares were identified by Roman numerals I - V for the rows and the letters A - H for the columns. A sample target page is shown in Figure 1. A point, to the nearest mile, was indicated to the subject by a pair of coordinates, such as C 8.4, II 7.3.

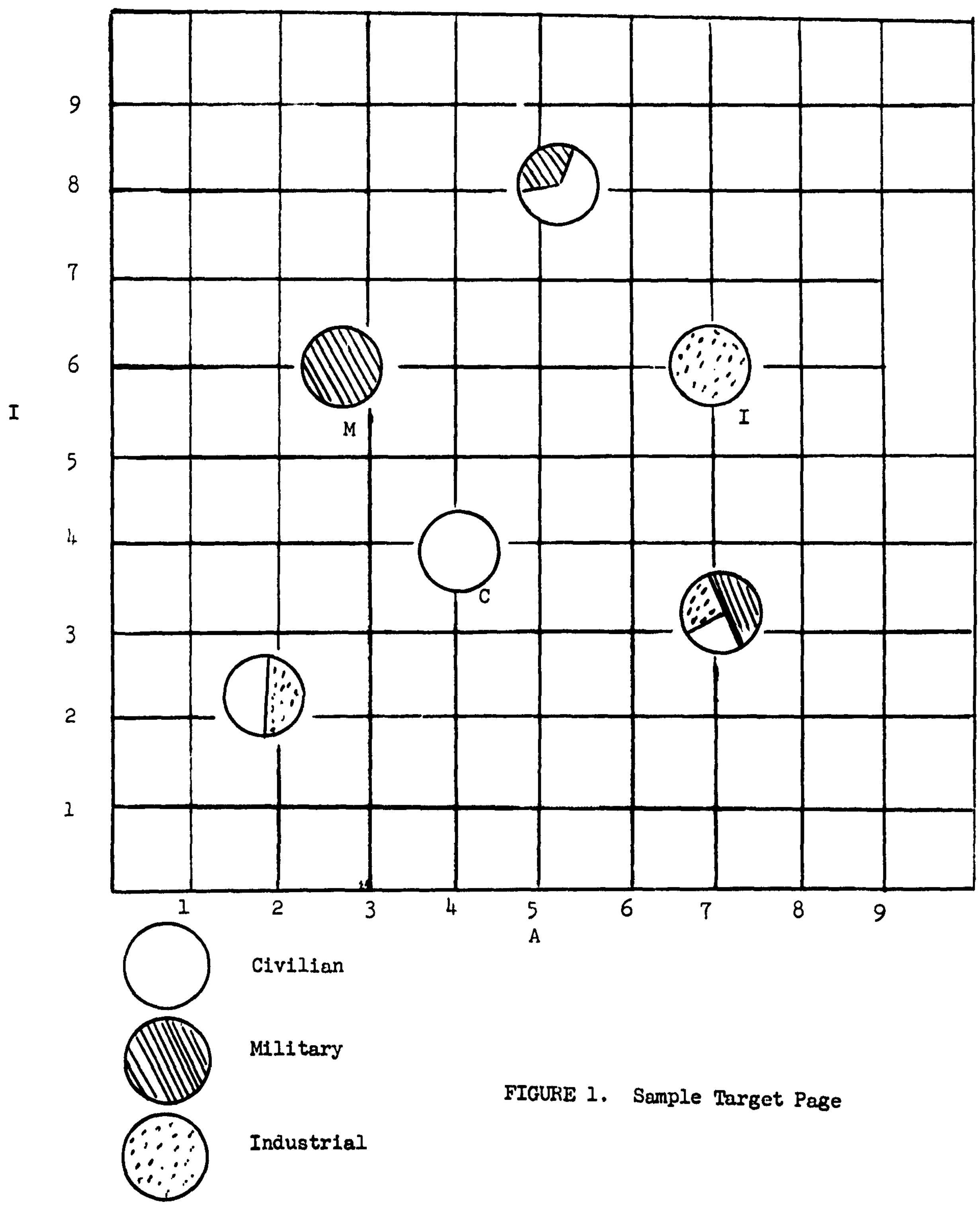


FIGURE 1. Sample Target Page

Targets were placed arbitrarily on each page of the target system. These targets were represented by a circle with a radius equivalent to 5 miles and were coded to indicate proportions within the target of the three major classes or resources with which the experiment was concerned. Each page contained three separate targets which were totally occupied by only one of the three classifications of resources: military. industrial, or civilian. Three other targets were arbitrarily placed on each page. These targets were made up of various proportions of the three classes. A key to the coding and an index to the area page appeared along with the georef grid on each page of the target booklet.

The Bombing Model

With the hypothetical target system in hand, the experimental stimulus data were generated. These data provided both the stimuli for the subjects and the "true" value of $P(D/H_i)$ against which the performance of the subjects were to be evaluated. The method was of the Monte Carlo type and a bi-variate normal distribution was taken to be characteristic of the bombing errors associated with the data. In applying the bombing model, each of the three classes of targets was taken in succession as the desired aiming point. The one target wholly associated with a given class was considered the desired ground zero on each of the forty pages representing the target system. Two deviations were randomly selected from a table of normal deviates[1], the z scores were translated into distance on the X and Y axes, and the resulting point was plotted as the datum for that case. If the distance obtained for either X or Y resulted in a point falling outside the georef grid, the two randomly selected z values were discarded and another pair was chosen from the tables.

1. RAND, A Million Random Digits with 100,000 Normal Deviates, (Glencoe, Illinois: The Free Press, 1955).

The translation to distance depended on the CEP which the model was using, and data were generated for three CEP's: 10, 15, and 20 miles. Each page of the target booklet, therefore, had nine items of data associated with it: one for each of the three uniform targets for each of the three CEP's used in the model. The three target classes were used in the experimental situation to define content variable, while the CEP selected was associated with a difficulty variable on the assumption that the larger the CEP, the more variable the data, and thus the more difficult the subject's estimate of P(D/H).

Probability Measurement

Once the data points had been placed by the bombing model, it was necessary to arrive at one P(D/H) for each point. This was done by assuming this value to be equal to the probability that the bomb would have landed as far away as it did or farther for each case. Scales were drawn corresponding to each of the circular error probabilities used, and the P(D/H)'s were read off for each point by direct measurement on the worksheets used. The values determined in this way are shown for sample case in Table I. The rows in the table identify the experimental variables by content and difficulty. The columns labeled X and Y show the coordinate of each point in the target system and the remaining columns indicate the "true" P(H/D) values for each of the three hypotheses associated with that data point. The main diagonal cells for the last three columns of Table 1 give the probabilities obtained when the target for that particular row was selected at the aiming point in the bombing model. For this reason in the great majority of cases the main diagonal cells contain higher values than the off-diagonal cells. The X and Y values of Table 1 are the only data presented to the subjects.

TABLE I

Sample Work Sheet

		A	I			
		X	Y	P(D/M)	P(D/I)	P(D/C)
CEP = 10	M	A3.1	I6.6	.70	.01	.01
	I	A5.5	I6.1	.01	.20	.01
	C	A4.3	I3.4	.01	.01	.85
CEP = 15	M	A3.1	I5.1	.75	.01	.45
	I	A7.7	I6.2	.01	.85	.01
	C	A3.8	I3.1	.05	.01	.82
CEP = 20	M	A2.1	I6.4	.90	.02	.16
	I	A5.5	I4.7	.22	.50	.65
	C	A4.9	I3.1	.09	.11	.81

Test Booklets

The test booklets were 8 1/2 x 11 inch size, three-ring binders. Each of the forty target map pages were enclosed in a glassene page protector, so that the booklets would remain relatively free from extraneous markings through continued use. The map pages were printed so that the long dimension formed the horizontal axis, and the booklets were handled, therefore, by placing the binding parallel to the front edge of the work surface and turning the pages away from the subject. Into the back side of each of the pages (the first page in the booklet was blank on the front) was inserted a 3" x 5" card with the coordinates of the bomb burst. The stimulus material for each probability estimate required of the subjects, was, therefore, the map page together with the coordinate designation facing it.

Independent and Dependent Variables

The experiment was concerned with three independent variables, referred to as condition, content, and difficulty.

1. There were two conditions studied in this experiment, PIP (P) and NON-PIP (N). In the PIP condition the subject was required to give estimates of the conditional probability of the data, given each of the hypotheses under consideration. He was required, in other words, to supply the term $P(D/H_i)$ for each H, so that the Bayesian Processor could be operated. In the N condition, subjects estimated P(H/D) directly, that is, after each item of the data was reviewed, a revised opinion of the probability of each hypotheses was given as the response.
2. The "content" variable was introduced not so much as an experimental variable as for convenience in design and as a way of controlling any bias that might have existed with regard to the names attached to the stimulus

elements. "Content" referred to that one of the three hypotheses which was "true" in a particular case, i.e., whether Military, Industrial, or Civilian targets was the class at which the enemy was aiming.

3. The "difficulty" of a stimulus set was varied by the selection of the CEP for the model when the bomb burst points were generated. CEP's of 10, 15, and 20 miles were selected, with difficulty assumed to increase as the number became larger, since the burst tended to be less clustered about the targets at which they were aimed.

Any one set of stimulus material was defined by its three designations, for example, M-10-P. This meant that Military was the true strategy of the enemy, his weapons had a CEP of 10 miles, (the lowest degree of difficulty) and the subjects were performing in the PIP condition of the experiment.

The dependent variable in the experiment has an S's estimate of the relevant probabilities, $P(D/H_i)$ for the PIP condition and $P(H_i/D)$ for the NON-PIP condition. In each case, three probability estimates were required, and the S's response was recorded on a response sheet, an example of which is shown in Figure 2. S placed a mark on each scale and beside it wrote the number representing his opinion. His mark on the scale labeled M, represented, in the PIP case, for example, his estimate o' ⁻(D/H_m), and the number beside the mark was used as the input to the Bayesian p ⁻essor.

Subjects

The subjects for the experiment were undergraduate students from an introductory psychology class, averaging in age from 18-24. Two of the subjects were female. The subjects were paid on a straight-time basis, and none had previously participated in an experiment of this type. All the subjects were considered naive with respect to their understanding of the probability concepts involved in the experiment.

Subject______________ Date______________ Item____________

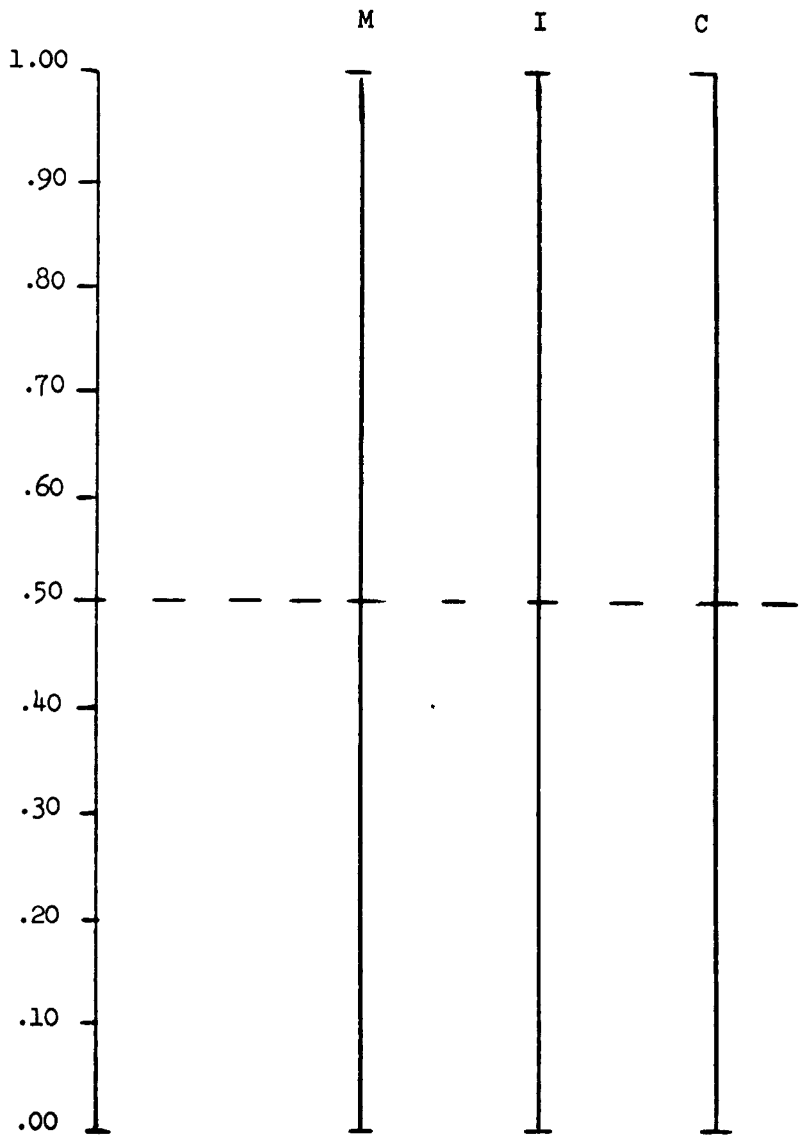


FIGURE 2

Sample Response Sheet

Experimental Design

Each of the 18 subjects was required to participate in six experimental sessions, three for each of the P and N conditions. Each received three different sets of stimulus materials. The same materials were used for the P and N conditions for any one subject. The subjects were divided into three groups, so that each received a particular combination of the content and difficulty variables. Each subject performed once, therefore, on each of the content and difficulty variables, for each condition. The order of running was randomized, with the restriction that no subject was to receive the same content variable twice in a row nor was he to receive the same condition three times in a row.

The experimental design called for all subjects to take six of the possible eighteen treatment combinations. It followed, as in the Lindquist Type VII Design (Lindquist, 1953). Table II summarizes this design.

TABLE II

Experimental Design

	P			N		
	M	I	C	M	I	C
Group 1	10	15	20	10	15	20
Group 2	15	20	10	15	20	10
Group 3	20	10	15	20	10	15

Instructions:

The subjects were assembled in one room and given a "lecture" about the nature of the experiment by one of the E's. On the chalk board was drawn one sample page from one of the "test booklets" they would be using during the experiment. This drawing was ..erred to during the lecture. The E also had a sample booklet and a sample response sheet, which he used in describing the task to the subjects. The text of the instructions are presented in Appendix A. They have been summarized below.

The subjects were told to imagine themselves as military commanders receiving NUDETS reports. They were to use this report to locate the bomb blast on the map section and then make probability judgments about the bomb blast. In one type of judgment (Judgment Type I), which was the NON-PIP condition, they were told to make three judgments estimating the probability that the enemy was aiming at Military, Industrial, or Civilian targets, given the information that the bomb blast occurred where it did. This was equivalent to asking them to estimate the posterior conditional probability P(H/D) directly.

They were told that an answer of 1.0 meant that they had complete certainty that the enemy was using a particular strategy. An answer of 0 meant they had complete confidence the enemy was not using a particular strategy. They were told, however, that answers of 1.0 or 0 should never be used because the data could never be accurate enough to allow them that degree of certainty. They were told to assign numbers of .01 to .99 to indicate the amount of certainty they had in their judgments. The subjects were also told that they did not have to normalize their estimates, i.e., make them sum to one, but if they assigned a high number to one of the targets, then,

they should assign low numbers to the other two targets. Furthermore, they were urged to "accumulate" their knowledge in Judgment Type I, i.e., they should use their prior experience in giving their probability estimates.

In Judgment Type II (PIP condition), the subjects were told to assume each hypothesis in turn as being true and to estimate the probability that the bomb blast occurred where it did. This was equivalent to having them estimate the conditional probability of the datum (geographical location of the bomb blast) given the hypotheses, i.e., P(D/H). They were told to assign numbers between .01 and .99 to indicate the amount of certainty they had in their judgment. They were also cautioned that each judgment should be kept completely independent of all other judgments they have made. They were to consider each datum as an independent entity. They were also explicitely told not to normalize their judgments. It was pointed out to them that when they were making Judgment Type II, the three numbers they gave as their probability estimates for each page of the test booklet could be all high, all low, or any combination of highs and lows.

The subjects made both types of judgments during the experiment but only one type of judgment at a time, i.e., for each test booklet. To assure that the subjects knew what judgment they were to be making, a card describing the type of judgment was placed before them. These cards contained the following statements respectively:

JUDGMENT TYPE I

LOCATE THE BOMB BLAST ON THE MAP SECTION

ESTIMATE THE PROBABILITY THAT THE ENEMY IS AIMING AT EACH OF THE THREE TARGETS, MILITARY, INDUSTRIAL, OR CIVILIAN.

JUDGMENT TYPE II

LOCATE THE BOMB BLAST ON THE MAP SECTION

ASSUME THE ENEMY HAS AIMED THE BOMB AT EACH OF THE TARGETS (MILITARY, INDUSTRIAL, OR CIVILIAN). THEN ESTIMATE THE PROBABILITY THAT THE BOMB BLAST OCCURRED WHERE IT DID.

Experimental Procedure

The eighteen subjects who participated in the experiment made all their judgments in one day. The instructions were given to them, and questions about the experiment were answered. This procedure took approximately one and a half hours, and the experimental sessions commenced immediately after.

Each subject received his stimulus material and response sheet booklet for the first session, together with a card to be placed in front of him as a constant reminder of the type of judgment (I, or II, which corresponded to NON-PIP and PIP respectively) that he was to be making during the session. As each subject finished the booklets for a particular session, he was instructed to bring them to the experimenter and wait until the next set of material was given to him. The experimental material consisted of eighteen booklets, one for each session; for the session to follow, the entire set of booklets had to be shifted to different subjects. As a result, some subjects were left without material on which to work, while others had booklets waiting for them on their desks. In no case did a subject have to wait for more than thirty minutes.

The subjects sat at desks measuring about 3 x 4 feet arranged in two rows of nine each in a single room. One of the three experimenters was present in the room at all times to answer questions. The experimental sessions started at 10:30 a.m., and the last subject had finished by 4:00 p.m. A 30-minute break for lunch was allowed at

noon, at which time all S's stopped wherever they were and continued from that point after the lunch period. The experimental atmosphere was informal; S's were allowed to smoke and to leave their desks or the room as they desired. Talking among S's, especially during periods where two or more were waiting between sessions, was, however, discouraged. When a subject brought in his last set of materials, he was paid for his time at the rate of $1.50 per hour.

RESULTS

Bayesian Processor

As described in a previous section, the $P(D/H_i)$'s generated by the bombing model may be considered as the "true" or "ideal" conditional probabilities. These "ideal" probabilities were fed into the Bayesian Processor programmed on the AN/FSQ-32 computer for the nine experimental PIP conditions, and the posterior conditional probabilities P(H/D) were calculated. In all except those cases with the highest variability (CEP = 20), the P(H/D) reached the maximum (P = .98) on the second trial. For the conditions with high variability (CEP = 20), the P(H/D) reached a maximum in four trials. Thus from an ideal standpoint these data were very "easy". The generation of the datum points using the bombing model yielded early (i.e. within the first few trials) likelihood ratios ranging from 60:1 to 90:1. These likelihood ratios, along with the fact there were few reversals during the course of the 40 trials, resulted in an early rise to a maximum P(H/D), which did not appreciably drop below maximum for the duration of the trials.

Likelihood Ratio Tests

The very high likelihood ratios that the bombing model yielded for this study are unrealistic although experimentally convenient. To get a better idea of the

sensitivity of Bayes's theorem to likelihood ratio, a series of constant likelihood ratios from 1.05 (relatively low) to 3.00 (relatively high) were used to see what their effects would be on the number of trials to reach maximum. The results are presented in Table III.

In this Table, the numbers represent the probability of the correct hypothesis. For the low likelihood ratio (1.05), the highest P(H/D) after 40 trials was only .80. As may be seen in Table III, as the likelihood ratio increased, the number of trials to maximum decreases rapidly; and for a likelihood ratio of 3.00, the maximum P(H/D) = .98 is reached after only three trials. It may be seen more clearly, now, why it was said that the bombing model generated "easy" data. The likelihood ratios for the data used in the experiment were 20-30 times greater than 3.00.

Empirical Results

The basic response variables of this study were the output of the Bayesian Processor (posterior probability) for each trial in the PIP condition and the posterior probability that was provided directly from the subjects for each trial in the NON-PIP condition. These posterior probabilities were normalized in another computer program.

The results of the experiment show that the highest probability was always assigned to the correct hypothesis, i.e. whether the true target was Military, Industrial, or Civilian. The means for the forty trials for experimental conditions are shown in Table IV. The forty trials were also divided into eight blocks of 5 trials each, and the mean posterior probability for each block was calculated. The results are plotted in Figure 3. As may be seen in Table IV and Figure 3, the PIP condition was superior to NON-PIP in (a) reaching higher levels of posterior probability, and (b) reaching asymptote at a faster rate in all cases except one (M-10). Even though the results

TABLE III

BAYESIAN PROCESSOR TESTS (Constant Likelihood Ratio [L.R.])

L.R.	VALUE OF P(HID) AFTER TRIAL NUMBER: 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1.05	35	38	37	38	40	41	42	43	45	46	47	49	50	51	52	54	55	56	58	59
1.10	35	38	40	42	44	47	49	51	54	56	58	61	63	65	67	69	71	73	75	76
1.15	37	40	43	47	50	54	57	61	64	67	70	73	76	78	81	83	85	86	88	90
1.20	38	42	46	51	55	60	64	68	72	75	78	81	84	86	88	90	91	93	94	95
1.25	39	44	50	55	61	66	71	76	80	83	86	88	91	92	94	95	96	97	97	98
1.30	40	46	53	59	65	71	76	81	84	88	90	92	94	95	96	97	98			
1.40	41	50	58	70	73	79	84	88	91	94	95	97	98							
1.50	43	53	63	72	79	85	90	93	95	97	98									
1.60	44	56	67	77	84	89	93	96	97	98										
1.70	46	59	71	81	88	92	96	97	98											
1.80	48	62	75	84	91	95	97	98												
1.90	49	64	78	87	93	96	98													
2.00	50	67	80	89	94	97	98													
2.25	53	72	85	93	97	98														
2.50	57	60	89	95	98															
2.75	58	79	91	97	98															
3.00	60	82	93	98																

L.R.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1.05	60	61	63	64	65	66	67	69	70	71	72	73	74	75	76	77	78	79	80	80
1.10	78	80	81	83	84	85	86	87	88	89	90	91	92	92	93	94	94	95	95	96
1.15	91	92	93	94	95	95	96	96	77	97	98									
1.20	96	96	97	97	98															

TABLE IV.

MEAN POSTERIOR PROBABILITY FOR THE EXPERIMENTAL CONDITION

Hypothesis / CEP	PIP M	PIP I	PIP C	PIP Grand Mean	NON-PIP M	NON-PIP I	NON-PIP C	NON-PIP Grand Mean
10	94.9	94.9	93.5	94.4	80.8	69.5	63.8	71.3
15	92.8	93.2	95.0	93.7	64.8	77.0	59.5	67.1
20	93.4	90.4	94.6	92.9	65.4	55.8	68.6	63.2
Grand Mean	93.8	92.8	94.4		70.3	67.4	64.0	

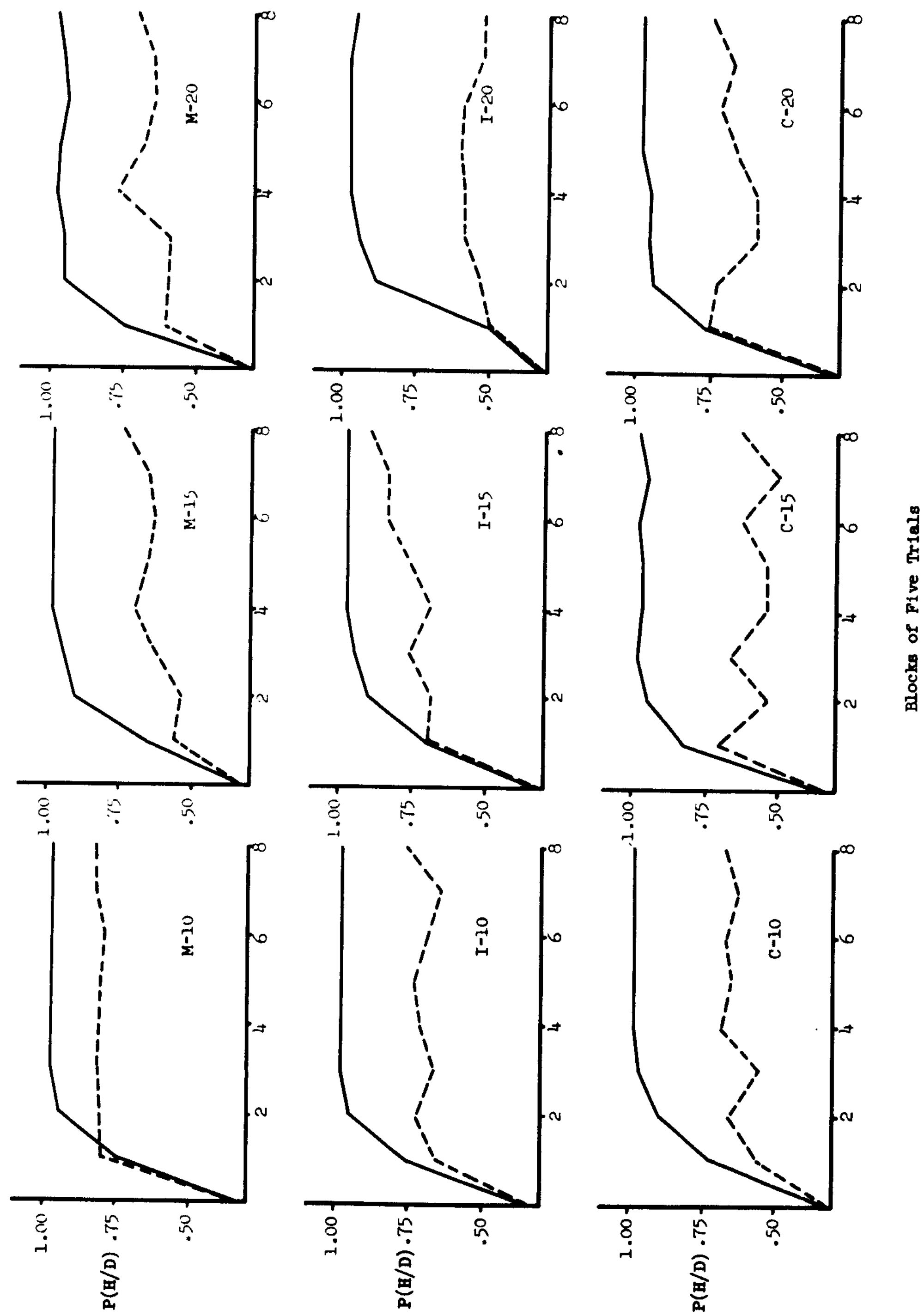


FIGURE 3. MEAN POSTERIOR PROBABILITY FOR PIP AND NON-PIP

of the experiment seemed immediately apparent, an analysis of variance was performed on the mean responses for the first 20 trials. The first 20 trials were chosen because in almost all cases, the mean posterior probability was approximately asymptotic by the end of 20 trials. The summary of this analysis is presented in Table V. Clearly, the difference between the two major experimental conditions

TABLE V.

Summary ANOVA Table:

Source			of	SS	MS	F
Between Subjects			(17)	5994		
Content x Difficulty	(b)		2	1475	738	2.45
error	(b)		15	4519	301	
Within Subjects			(90)	24183	269	
Content	(C)	1	2	167	84	3.18
Difficulty	(D)	1	2	371	186	7.04 *
Experimental Condition	(E)	3	1	16109	16109	217.39 **
Content x Difficulty	(w)	1	2	32	16	-
Content x Experimental Condition		2	2	415	208	1.50
Difficulty x Experimental Condition		2	2	551	276	1.99
Content x Difficulty	(b) E	3	2	500	260	3.51
Content x Difficulty	(w) E	2	2	951	476	3.43
error	(w)		(75)	5067	67.6	
$error_1$	(w)		30	760	26.4	
$error_2$	(w)		15	2083	138.9	
$error_3$	(w)		30	2224	74.1	
Total			107	30177		

(PIP vs NON-PIP) is highly significant. The difficulty variable was also significant, but inspection of Table IV indicates that difficulty influenced NON-PIP (increasing difficulty lowered performance) but did not seem to effect PIP. All other main effects and interactions were not significant.

Scatter Plots

One other empirical result is worth mentioning. In performing under the PIP conditions, the subjects were highly consistent in their probability judgments. They tended to give relatively low estimates when the "true" conditional probabilities were low and relatively high estimates when the "true"conditional probabilities were high. This is illustrated in a representative scatter plot in Figure 4.

Conclusions and Discussion

This was a pilot study to test the application of some principles of PIP theory outlined by Edwards (1962). If we consider the estimation of posterior probabilities directly by humans, and compare this with the posterior probabilities generated by Bayes's theorem using only an input from humans of the form P(D/H), there is no question as to the superiority of PIP. This study also supports the findings of an unpublished study by Hayes, Phillip, and Edwards at the University of Michigan, that humans seem to be quite conservative in coming to conclusions about which hypothesis is correct. The estimation of posterior probabilities in the NON-PIP condition are always much lower than they should be. In part, this result is largely an artifact of the normalizing process. When subjects do not normalize their estimate, i.e., make them sum to one, and these estimates are then automatically normalized, the normalization by its very nature results in lower posterior probabilities. There is another more serious difficulty here, however; the subjects in this experiment were not given a good opportunity to express the "real" degree of certainty they might have had in their estimates. On many occasions the subjects expressed the fact that they "were quite sure," which was the correct hypothesis but

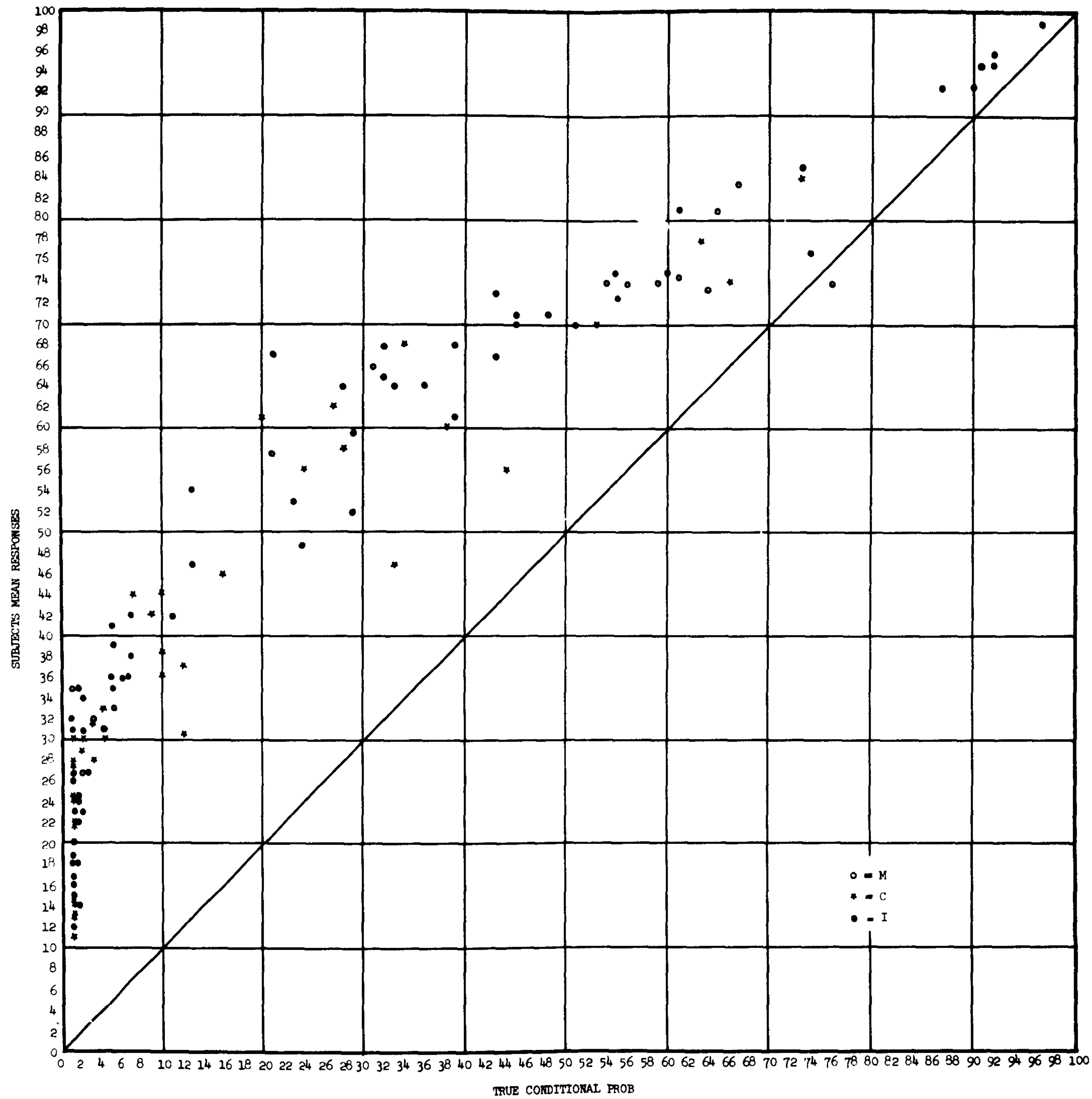


Figure 4. Scatter Plot for Condition M-20-P

then assigned a relatively low number (e.g., .6 to .7) as their estimate of P(H/D) for that hypothesis. There is no way to be certain, but it seems intuitively plausible that the mere assignment of numbers between .01 and .99 is not an effective way for humans to express the degree of confidence or certainty they have in their judgments. What is really needed is to have people make actual decisions or engage in behavior that actually commits them to a position. More will be said on this point later.

When human probability estimation in PIP is considered (i.e., estimating P(D/H), which is the inserted into Bayes's theorem and the resulting P(H/D) is then calculated), it makes little difference whether the estimates are conservative or liberal. All that is necessary is that the behavior in making estimates be reasonably consistent. By reasonably consistent we mean the ratios of the estimates (likelihood ratios) must be in the right direction, i.e., must favor the correct hypothesis. Bayes's theorem does all the rest. It is a powerful inferential tool that capitalizes upon whatever consistency it finds in the data and that rapidly achieves maximum (close to one) or minimum (close to zero) posterior probability values for the hypothesis that is correct or incorrect.

There were several deficiencies in this experiment. Some of these have been mentioned already, but it would perhaps be useful to summarize them here. It is expected that these shortcomings will be removed in the next study.

- The tasks for the subjects were too simple, too long, and too boring.
- Because of a lack of facilities it was not possible to give feedback to the subjects. In any application of PIP that would be at all meaningful, feedback would most certainly be present.

- Probability estimation by itself is not an interesting psychological or system problem; although the methodology of the psychophysics of probability is essential for future research on PIP. Combining probability estimation with actual decision making of the diagnostic and action-selection variety is of central interest in these studies. Perhaps the best way to do this would be to intordure costs and payoffs into the task situation, thus forcing the subjects into "commitment-type behavior."
- The technique that was used to generate the data, while adequate for this study, left much to be desired. It was not possible to have precise control over the characteristics of the data that were generated by a Monte Carlo method. Possibly the data points may be generated on a more controlled basis, one that will enable more precise specification of the difficulty and complexity of the data.

References

Edwards, W., "Dynamic Decision Theory and Probabilistic Information Processing," Human Factors, April 1962, pp 59-73

Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education, Houghton Mifflin Co., 1953.

APPENDIX A

Instructions

We are trying to develop techniques that will help people use data or information in making certain decisions. Our major concern is with military decision making. We would like you to help us in this by making certain judgments in a task that will be explained to you shortly.

There are two basic judgments we want you to make. To gain an intuitive feel for what they are like and how they differ from each other, consider the following decision making situation. Suppose you had to decide or predict that it was or was not going to rain tomorrow; that is, there are only two possible conditions, rain or no rain, and you have to make a judgment about which it will be. You, of course, are never perfect at this and can only make a "probable" judgment. For example, you could use whatever information you had available to you and decide that there is a 50-50 chance of rain tomorrow, or an 80% chance of rain tomorrow, etc. This is a very common type of judgment, and weather forecasters do it all the time. We will want you to make such judgments in the problem we are going to give you.

We will also want you to make another type of judgment, however, that differs from what was just described. In this type of judgment you do not try to predict whether or not it will rain tomorrow. Instead, you assume that it will rain tomorrow; that is, consider this to be absolutely true. With this in mind you then make probability judgments about events or data that you observe around you. For example, you observe it is a very cloudy day. Now, what is the probability of it being very cloudy given

that it will indeed rain tomorrow? You then entertain the only other possible condition, i.e., it will not rain tomorrow. Now, what is the probability that it is a very cloudy day given that it will not rain tomorrow? When you make judgments like this, we can convert the information you give us into an estimate of whether it will rain tomorrow or not using a probabilistic formula called Bayes's theorem. (The details of the formula are not important. The proper use of this formula requires you to make probability judgments of the kind "what is the likelihood of observing an event, given that some other condition is true." The formula then calculates the probability of the condition, given the event. One of the main objectives of the experiment, in which we are asking you to participate, is to discover if the use of Bayes's formula works as well or better in the decision task than if we were to ask you directly what conditions are going to exist. In the task we want you to perform, it is extremely important that you keep in mind the type of probability judgment you are making. For convenience we call them Judgment Type I and Judgment Type II. We will tell you each time which type you are being asked to make. Now, we will describe the task to you.

Imagine that an enemy is attacking the United States by firing a large number of small (low yield) nuclear bombs at selected targets across the nation. Each bomb, as it lands and detonates, is detected by a system called NUDETS for Nuclear Detonations. The information that NUDETS gives that you will be concerned with is the geographic location of the bomb blast. The information is reported to you and looks like this (3" x 5" card on example page of booklet).

A 4.6

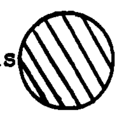
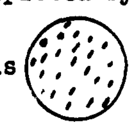
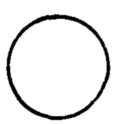
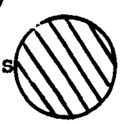
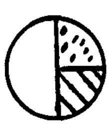
II 7.8

The first line is always a letter from A to H, followed by a number from 0.1 to 9.9. The second line is always a Roman numeral from I to V, followed by a number from 0.1 to 9.9. This information locates the bomb blast on a map of the target area. This map has been divided into 40 smaller sections labeled AI, AII, etc., up to HV (Grid on sample page).

To locate the bomb blast, use the first line of the bomb report and use A 4.6 to locate the bomb relative to the horizontal axis, A. The second line, I 7.8 locates the bomb relative to the vertical axis, I. Thus, this bomb fell right <u>here</u> (sample page on chalk board).

Do you understand how to locate the bomb blast?

There are 40 pages similar to this one in this book. Each page is a 100-mile-square section of a map. Thus each small block within the section is a ten mile square.

The map sections also show the location of targets at which the enemy might be aiming. There are three major types of targets: Military, Industrial, and Civilian. On the map section a Military target is depicted by lines, like this . An Industrial target is depicted by dots like this . A Civilian target is a plain circle like this . A target coded like this  means that it has 100% Military importance, with no Industrial and no Civilian importance. A target coded like this  means that the target has 50% Civilian importance, 25% Military importance, and 25% Industrial importance. (All examples were drawn on the chalk board.)

An example of a pure Military target would be a United States Air Force missile base. A pure Industrial target might, for example, be an automobile factory. A pure Civilian target might be a residential section of a city, etc. An example of a combination target, that is, for example, 50% Military and 50% Industrial might be a Manufacturing factory that turns out both automobiles and missiles. Are there any questions about the coding of these targets?

The enemy has 40 bombs and shoots one of them at each of the 40 different geographical areas of the country that are represented by each map section in this book. One bomb is aimed at a specific target on each map section.

The enemy is operating under one and only one of three strategies: 1) he aims all his bombs at Military targets, or 2) he aims all his bombs at Civilian targets, or 3) he aims all his bombs at Industrial targets. Suppose the enemy is aiming at Military targets. This means that he aims one bomb at one of the targets shown on each page. He aims at that target on each page which has the highest Military importance. If he is shooting at Civilian targets, he aims all 40 bombs at the target on each page with the highest Civilian importance. The enemy never uses a mixed strategy, aiming some bombs at Civilian targets and some bombs at Industrial or Military targets.

The overall problem of this task is to decide which of the three strategies the enemy is using; that is, is he aiming at Civilian, Industrial, or Military targets?

If the enemy's aim were perfect, it would be easy to tell what he is aiming at. But the enemy's aim is not perfect; the bombs do not always fall where he aims them. The aiming errors of the enemy's missiles are crucial to this task; unfortunately, we cannot tell you very much about them. We can tell you that this is quite realistic.

Almost all data of the kind we are talking about is replete with error and therefore quite fallible.

We can also tell you that the error has no bias. That is, the bomb is just as likely to fall 10 miles North of the target as it is to fall 10 miles South of the target, or 10 miles in any direction away from the target.

Also, the bomb is more likely to fall close to the target than far away from it. That is, if the enemy is aiming at one of the targets, usually the bomb will land near that target. By "usually" we mean most of the time.

We mentioned before that the eventual aim of this task is to decide what type of targets (Military, Industrial, or Civilian) the enemy is attacking. We also mentioned that we want you to make two quite different judgments but use only one type of judgment at any one time. In Judgment Type I, we want you to assume you are a military commander receiving NUDETS reports. You use the report to locate the bomb blast on the map section, and then give a probability statement of what strategy the enemy is using; that is, report whether he is aiming at Military, Civilian, or Industrial targets. For each map section (booklet page) you will make three judgments; the Probability that the enemy is aiming at Military targets, the Probability that he is aiming at Industrial targets, and the Probability he is aiming at Civilian targets. We want you to give us a number greater than .00 and less than 1.00 expressing the likelihood that the enemy is aiming at the respective targets.

An answer of 1.00 means that you have complete certainty that the enemy is using a particular strategy. An answer of .00 means you are absolutely sure the enemy is not using a particular strategy. From what we have already told you about these data you will never want to use the numbers .00 or 1.00. Because of error in the

data, you can never be completely sure about your judgments.

Do not use the response of one (1.00); also, do not use the response of zero (.00). You should be giving us numbers somewhere between .01 and .99. For example, the number .25 means there is about a 25 per cent chance that the enemy is aiming at one of the targets. Another way of thinking about a response of .25 is that there is one chance in four that the enemy is using a particular strategy.

Indicate your judgment on these sheets of paper by making a small mark on each of the three lines labeled M for Military, I for Industrial and C for Civilian. Also write what the number is beside your work, for example, if you think the probability is .25 that the enemy is aiming at one of the three possible targets, make a mark and write the number (show). You are to make these judgments for 40 different bombs, one for each page. Are there any questions about the type of judgment you are to make here?

In Judgment Type I it is not necessary for the sum of the number you assign to add up to any one number. However, in Judgment Type I, if you assign a high number to one of the targets indicating you are quite sure that is the target the enemy is shooting at, then you should assign low numbers to the other two targets. Also in Judgment Type I, you should use whatever experience you have gained in making your judgment. If you wish, you may turn bakc the pages and see what responses you have given before.

Now we will describe the other type of judgment, Judgment Type II, we want you to make. We are still trying to decide what type of target (Civilian, Industrial, or Military) the enemy is attacking. However, we do not want to get this information

directly from you. You are still to consider yourself as a military commander receiving NUDETS reports, and again you use the report to locate the bomb blast on the map section.

Now your task will be as follows: imagine that, in fact, the enemy is attacking one of three possible targets on the map section. Consider each one in turn and suppose the enemy has aimed the bomb at that target. Accept for a moment that this is <u>true</u>. Now, if the enemy has <u>really</u> aimed at one of the targets, what are the chances that the bomb would land right where it did? We want you to give us a number greater than zero and less than 1.00 that expresses the likelihood that if the enemy were indeed aiming at a particular target, a bomb would land where it did. An answer of .00 from you would mean: "When the enemy tries to bomb this target, one of their bombs couldn't possibly land <u>there</u>." An answer of 1.00 would mean: "When the enemy tries to bomb this target, their bomb will surely land exactly there." You can tell from this that you'll never want to use the number 1.00, because nobody, under any conditions, can be sure that a bomb will land in any one particular spot. Also you would never want to use .00 because there is always some possibility, however small, that the bomb could land there.

In Judgment Type II, we do <u>NOT</u> want you to tell us your guess as to whether or not the enemy is aiming at a particular target or has a particular strategy in mind. You must assume at once that they are aiming at that target, and then estimate how likely it is that the bomb would land where it did.

It is very important that each time you are presented with a NUDETS report you forget or disregard all previous information you have received, all previous responses you have given, and any hypothesis you may have formed about the enemy's intentions.

Think only of the problem at hand: Assuming the bomb is aimed at one of the three targets, how likely is it that the bomb would land where it did. Also it is possible in Judgment Type II for your judgments to be all high numbers, all low numbers or any combination of high or low numbers. This is true because you are considering each bomb blast independently for the three targets.

You are to make these judgments for 40 different bombs, one for each page. Indicate your judgment on the sheets of paper just as we described for you before by making a small mark on each of the three lines labeled M, I, and C and writing the number you mean beside it.

Are there any questions?

Remember, you will be making both types of judgments in this experiment, but at any one time you will be making only one type of Judgment. Each time we give you a book to work with, we will tell you what type of Judgment we want you to make. To help you remember, we will indicate on a card what type Judgment, I or II, we want you to make. You should keep this card in front of you.

After the instructions had been read aloud to the subjects, they were told that they could ask questions at any time during the experiment. When the subjects were given their second booklet out of the six they were to receive, they were asked by one of the experimenters if they understood the task and if they wanted more instruction. At this time the difference between Judgment Type I and II was repeated to them.

The subjects were taken to another room in which they worked at individual desks. They were told that when they finished each booklet to return it to the experimenter. They were then allowed to proceed at their own pace to the next booklet.

UNCLASSIFIED

System Development Corporation,
Santa Monica, California
A STUDY IN PROBABILISTIC INFORMATION
PROCESSING (PIP).
Scientific rept., TM-1150/000/00, by
R. J. Kaplan, J. R. Newman, 2 April 1963,
36p., 2 refs., 4 figs.
(SD-97, ARPA)

Unclassified report

DESCRIPTORS: Communication Theory.
Decision Making.

Presents a pilot study that tests the application of some principles of PIP

UNCLASSIFIED

UNCLASSIFIED

(Probabilistic Information Processing) theory outlined by Edwards (1962). Supports the findings of an unpublished study by Hayes, Phillip, and Edwards at the University of Michigan, that humans seem to be quite conservative in coming to conclusions about which hypothesis is correct. Summarizes several deficiencies in this experiment: 1) The tasks given the subjects were too simple, long and boring; 2) Because of a lack of facilities it was not possible to give feedback to the subjects: 3) Probability estimation by itself is not an interesting psychological or system problem; and 4) The technique used to generate the data, while adequate for this study, left much to be desired.

UNCLASSIFIED